



A joint design for functional data with application to scheduling ultrasound scans[☆]

So Young Park^a, Luo Xiao^{b,*}, Jayson D. Willbur^c, Ana-Maria Staicu^b,
N. L'ntshotsholé Jumbe^d

^a Eli Lilly and Company, Indianapolis, IN, USA

^b North Carolina State University, Raleigh, NC, USA

^c Metrum Research Group LLC, Tariffville, CT, USA

^d Bill & Melinda Gates Foundation, Seattle, WA, USA



ARTICLE INFO

Article history:

Received 9 May 2017

Received in revised form 12 January 2018

Accepted 13 January 2018

Available online 31 January 2018

Keywords:

Covariance function

Functional data analysis

Fetal growth

Longitudinal data

Prediction

ABSTRACT

A joint design for sampling functional data is proposed to achieve optimal prediction of both functional data and a scalar outcome. The motivating application is fetal growth, where the objective is to determine the optimal times to collect ultrasound measurements in order to recover fetal growth trajectories and to predict child birth outcomes. The joint design is formulated using an optimization criterion and implemented in a pilot study. Performance of the proposed design is evaluated via simulation study and application to fetal ultrasound data.

© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Functional data analysis has been a popular statistical research area for the last two decades and has found application in many fields such as brain imaging (Jiang et al., 2009; Greven et al., 2010; Reiss and Ogden, 2010; Lindquist, 2012; Lu and Marron, 2014; Park and Staicu, 2015), biosignals (Crainiceanu et al., 2012; Randolph et al., 2012; Goldsmith and Kitago, 2016), genetics (Tang and Müller, 2009; Reimherr and Nicolae, 2014) and wearable computing (Morris et al., 2006; Li et al., 2014; Xiao et al., 2015). For a comprehensive treatment of functional data analysis see Ramsay and Silverman (2002, 2005) and Horváth and Kokoszka (2012).

This paper considers sampling design for noisy growth data. The motivation arises from the study of fetal growth, where measurements of fetal size may be obtained during pregnancy using ultrasound. And the particular question to be addressed is: when a fixed number of ultrasound scans will be taken during pregnancy, what are the optimal time points for data collection? Optimality can be defined either in terms of recovering individual fetal growth trajectories or in terms of predicting a birth outcome, such as birth weight. However, in practice it may be important to predict both individual growth trajectories and birth outcomes, and in such cases a joint optimality criterion must be formulated. We also consider the closely related question of the number of ultrasound scans required to achieve a desired level of optimality.

We address this question within the functional data framework. Design for functional data has received some interest recently. For example, Ferraty et al. (2010) considered a nonparametric model with a scalar response and a functional

[☆] Supplementary materials are available with this article at the *Computational Statistics and Data Analysis* website.

* Correspondence to: Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.
E-mail address: lxiao5@ncsu.edu (L. Xiao).

predictor and [Delaigle et al. \(2012\)](#) studied a similar problem for classifying and clustering functional data. Both methods are restricted to densely sampled functional data and focus on dimensionality reduction for a dense functional predictor. And for spatially correlated functional data, [Rasekhi et al. \(2014\)](#) and [Bohorquez et al. \(2015\)](#) considered the problem of selecting spatial sampling points.

Design for functional data has also been extended to longitudinal data. [Ji and Müller \(2017\)](#) proposed prediction-based criteria for sampling functional data with the target of either recovering individual functions or predicting a scalar outcome. [Wu et al. \(2017\)](#) exploited the mixed effects model representation of functional data and proposed a design criterion based on Fisher's information matrix of eigenvalues of the covariance function. There are several limitations with these approaches. [Wu et al. \(2017\)](#) focused on recovering individual functions, while [Ji and Müller \(2017\)](#) were limited to the study of design separately and did not consider a joint design, which is the focus of our data application. In addition, in these works the number of design points was pre-fixed and no data-driven method was developed. Finally, [Ji and Müller \(2017\)](#) did not compare functional data models versus parametric mixed effects models for prediction-based designs. Our work addresses these gaps.

Following early work on design, such as [Ylvisaker \(1987\)](#) and the references therein and recent work by [Ji and Müller \(2017\)](#), we consider prediction-based designs and propose a unified design criterion for both recovering individual functions as well as predicting scalar outcomes from a functional predictor. We also propose a practical data-driven method for selecting the number of design points, building on the result that the larger the number of design points, the better the prediction will be (see [Theorem 1](#)). Finally we conduct a comprehensive simulation study to evaluate the performance of functional data models as compared to parametric mixed effects models, and demonstrate numerically that functional data models might be preferred over parametric mixed effects models for prediction-based optimal designs for longitudinal data.

The rest of the paper is organized as follows. In [Section 2](#) we introduce functional data models and propose a unified prediction-based design criterion for sampling functional data. In [Section 3](#) we study the theoretic properties of the proposed design. In [Section 4](#) we discuss implementation of the design and propose a data-driven method for selecting the number of design points. In [Section 5](#) we illustrate the proposed method using a fetal ultrasound data. In [Section 6](#), we investigate the performance of the design via simulation studies.

2. Optimal design for functional data

In this section, we first describe functional data models and then formulate two optimal design problems for sampling functional data: one design targets accurate prediction of individual functions while the other targets accurate prediction of a scalar outcome. Then, we propose a unified design criterion that targets both recovering individual functions and predicting a scalar outcome. In particular, the unified design contains the previous two designs as special cases.

2.1. Statistical models

Consider a random function $X(t)(t \in \mathcal{T})$ defined over a continuous and compact time domain \mathcal{T} . Suppose that $X(\cdot)$ is a Gaussian process with mean function $\mu(t) = E\{X(t)\}$ and covariance function $r(s, t) = \text{Cov}\{X(s), X(t)\}$. We assume that $X(\cdot)$ is square integrable in \mathcal{T} and without loss of generality we let $\mathcal{T} = [0, 1]$.

In practice, $X(\cdot)$ is observed at a finite number of time points and contaminated with noise. Hence, for a random function $X_i(\cdot)$ with a subject index i observed at p time points $(t_1, \dots, t_p)' \in \mathcal{T}^p$, the observations are

$$W_{ij} = X_i(t_j) + \epsilon_{ij}, \quad 1 \leq j \leq p, \quad (1)$$

where the ϵ_{ij} are i.i.d. $\mathcal{N}(0, \sigma_\epsilon^2)$ and independent of $X_i(\cdot)$.

Let Y be a scalar outcome with a functional predictor $X(\cdot)$. And consider the functional linear model

$$Y = \alpha + \int_{\mathcal{T}} \bar{X}(t)\beta(t)dt + e, \quad (2)$$

where α is an intercept, $\bar{X}(t) = X(t) - \mu(t)$, $\beta(t)$ is a smooth coefficient function, and e is white noise independent of $X(\cdot)$ with mean zero and variance σ_e^2 .

The fundamental element in functional data analysis is the covariance function $r(s, t)$. By Mercer's theorem, $r(s, t)$ can be written as $\sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(s)\phi_\ell(t)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ is the collection of eigenvalues and the $\phi_\ell(\cdot)$ are the associated eigenfunctions which satisfy $\int_{\mathcal{T}} \phi_\ell(t)\phi_{\ell'}(t)dt = 1_{\{\ell=\ell'\}}$. Here $1_{\{\cdot\}}$ is 1 if the condition inside the bracket holds and 0 otherwise. To ensure that $\beta(t)$ is identifiable, we assume that the coefficient function $\beta(t)$ can be written as $\sum_{\ell=1}^K \beta_\ell \phi_\ell(t)$, where the β_ℓ are scalars and, a possibly infinite K represents the number of non-zero eigenvalues.

2.2. Optimal design for predicting functions

Fix $p \geq 1$ and assume that p observations will be collected from a new subject. The goal is to select the p optimal sampling points in \mathcal{T} for predicting the new subject's curve with the smallest possible error.

Let $\mathbf{t} = (t_1, \dots, t_p)' \in \mathcal{T}^p$ be the vector of sampling points and $\mathbf{W}_{i^*}(\mathbf{t}) = \{W_{i^*}(t_1), \dots, W_{i^*}(t_p)\}'$ be the noisy observations for a new subject i^* . Under model (1), the best predictor of $X_{i^*}(t)$ conditional on $\mathbf{W}_{i^*}(\mathbf{t})$ is the best linear unbiased predictor (BLUP) of $X_{i^*}(t)$,

$$E\{X_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\} = \mu(t) + \mathbf{r}(t, \mathbf{t})'\Sigma_{\mathbf{W}}(\mathbf{t})^{-1}\mathbf{W}_{i^*}(\mathbf{t}), \tag{3}$$

where $\mathbf{r}(t, \mathbf{t}) = \{r(t, t_1), \dots, r(t, t_p)\}' \in \mathbb{R}^p$ and $\Sigma_{\mathbf{W}}(\mathbf{t}) = \text{Cov}\{\mathbf{W}_{i^*}(\mathbf{t})\} \in \mathbb{R}^{p \times p}$. For simplicity, we suppress the notation \mathbf{t} from $\Sigma_{\mathbf{W}}(\mathbf{t})$ and use $\Sigma_{\mathbf{W}}$. The optimal sampling points \mathbf{t} can be selected by minimizing the mean integrated squared error of the BLUP,

$$\mathcal{M}_1(\mathbf{t}) := E \int_{\mathcal{T}} [X_{i^*}(t) - E\{X_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\}]^2 dt. \tag{4}$$

The optimal design is then defined as $\mathbf{t}^{\text{opt}} := \arg \min_{\mathbf{t} \in \mathcal{T}^p} \mathcal{M}_1(\mathbf{t})$. And we simplify $\mathcal{M}_1(\mathbf{t})$ as

$$\begin{aligned} \mathcal{M}_1(\mathbf{t}) &= E \int \{\bar{X}_{i^*}(t)\}^2 dt - E \int [E\{\bar{X}_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\}]^2 dt \\ &= \int r(t, t) dt - \int \mathbf{r}(t, \mathbf{t})'\Sigma_{\mathbf{W}}^{-1}\mathbf{r}(t, \mathbf{t}) dt \\ &= \int r(t, t) dt - \text{tr}(\mathbf{R}\Sigma_{\mathbf{W}}^{-1}), \end{aligned}$$

where $\text{tr}(\cdot)$ is the trace operator and $\mathbf{R} = \int \mathbf{r}(t, \mathbf{t})\mathbf{r}(t, \mathbf{t}) dt \in \mathbb{R}^{p \times p}$ whose (j_1, j_2) element is given by $\int r(t, t_{j_1})r(t, t_{j_2}) dt = \sum_{\ell=1}^K \lambda_{\ell}^2 \phi_{\ell}(t_{j_1})\phi_{\ell}(t_{j_2})$. We write $\mathbf{R} = \Phi(\mathbf{t})\Lambda^2\Phi(\mathbf{t})'$, where $\Phi(\mathbf{t}) = [\phi_{\ell}(t_{j_1})]_{1 \leq j_1 \leq p, 1 \leq \ell \leq K} \in \mathbb{R}^{p \times K}$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K) \in \mathbb{R}^{K \times K}$. Note that $\int r(t, t) dt = \sum_{\ell=1}^K \lambda_{\ell} = \text{tr}(\Lambda)$. It is also easy to show that $\Sigma_{\mathbf{W}} = \Phi(\mathbf{t})\Lambda\Phi(\mathbf{t})' + \sigma_{\epsilon}^2\mathbf{I}_p$. Therefore,

$$\mathcal{M}_1(\mathbf{t}) = \text{tr}(\Lambda) - \text{tr} \left[\Phi(\mathbf{t})\Lambda^2\Phi(\mathbf{t})' \{ \Phi(\mathbf{t})\Lambda\Phi(\mathbf{t})' + \sigma_{\epsilon}^2\mathbf{I}_p \}^{-1} \right].$$

And if we let $\mathbf{S}(\mathbf{t}) = \Lambda\Phi(\mathbf{t})' \{ \Phi(\mathbf{t})\Lambda\Phi(\mathbf{t})' + \sigma_{\epsilon}^2\mathbf{I}_p \}^{-1} \Phi(\mathbf{t})\Lambda$, then we obtain the simplified form

$$\mathcal{M}_1(\mathbf{t}) = \text{tr}(\Lambda) - \text{tr} \{ \mathbf{S}(\mathbf{t}) \}. \tag{5}$$

2.3. Optimal design for predicting an outcome

Similar to Section 2.2, assume that p observations will be collected from a new subject indexed by i^* . But let the goal now be to select the p optimal sampling points in \mathcal{T} for predicting the new subject's scalar outcome with the smallest possible error.

Using the same notation as in Section 2.2, let $\mathbf{t} = (t_1, \dots, t_p)' \in \mathcal{T}^p$ be the vector of sampling points and $\mathbf{W}_{i^*}(\mathbf{t}) = \{W_{i^*}(t_1), \dots, W_{i^*}(t_p)\}'$ be the noisy observations for subject i^* . Under the functional linear model (2), $E(Y_{i^*}|X_{i^*}) = \alpha + \int_{\mathcal{T}} \bar{X}_{i^*}(t)\beta(t)dt$, where $\bar{X}_{i^*}(t) = X_{i^*}(t) - \mu(t)$. Then under the functional data model (1), the best predictor of Y_{i^*} conditional on $\mathbf{W}_{i^*}(\mathbf{t})$ is the best linear unbiased predictor of Y_{i^*} ,

$$E\{E(Y_{i^*}|X_{i^*})|\mathbf{W}_{i^*}(\mathbf{t})\} = \alpha + \int_{\mathcal{T}} E\{\bar{X}_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\} \beta(t) dt.$$

And the mean squared error for predicting $E(Y_{i^*}|X_{i^*})$ is

$$\mathcal{M}_2(\mathbf{t}) = E \left[\int_{\mathcal{T}} \bar{X}_{i^*}(t)\beta(t) dt - \int_{\mathcal{T}} E\{\bar{X}_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\} \beta(t) dt \right]^2.$$

Then the optimal design is $\mathbf{t}^{\text{opt}} := \arg \min_{\mathbf{t} \in \mathcal{T}^p} \mathcal{M}_2(\mathbf{t})$. Note that the mean squared error for predicting Y_{i^*} is $\mathcal{M}_2(\mathbf{t}) + \sigma_{\epsilon}^2$, which results in the same design. This design was studied in earlier work including Ritter (1996), and more recently Ji and Müller (2017).

By Eq. (3), $\int E\{\bar{X}_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\} \beta(t) dt = \left\{ \int \beta(t)\mathbf{r}(t, \mathbf{t}) dt \right\}' \Sigma_{\mathbf{W}}^{-1} \mathbf{W}_{i^*}(\mathbf{t})$. Thus,

$$\begin{aligned} \mathcal{M}_2(\mathbf{t}) &= E \left\{ \int \bar{X}_{i^*}(t)\beta(t) dt \right\}^2 - E \left[\left\{ \int \beta(t)\mathbf{r}(t, \mathbf{t}) dt \right\}' \Sigma_{\mathbf{W}}^{-1} \mathbf{W}_{i^*}(\mathbf{t}) \right]^2 \\ &= \iint r(s, t)\beta(s)\beta(t) ds dt - \left\{ \int \beta(t)\mathbf{r}(t, \mathbf{t}) dt \right\}' \Sigma_{\mathbf{W}}^{-1} \left\{ \int \beta(t)\mathbf{r}(t, \mathbf{t}) dt \right\}. \end{aligned}$$

Then, $\iint r(s, t)\beta(s)\beta(t) ds dt = \sum_{\ell=1}^K \lambda_{\ell} \beta_{\ell}^2 = \text{tr}(\boldsymbol{\beta}\boldsymbol{\beta}'\Lambda)$, where $\boldsymbol{\beta}$ is a K -dimensional vector with $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$. We also obtain $\int \beta(t)\mathbf{r}(t, t_j) dt = \sum_{\ell=1}^K \lambda_{\ell} \beta_{\ell} \phi_{\ell}(t_j)$ which leads to $\int \beta(t)\mathbf{r}(t, \mathbf{t}) dt = \Phi(\mathbf{t})\Lambda\boldsymbol{\beta}$. Therefore, we obtain the simplified form

$$\mathcal{M}_2(\mathbf{t}) = \text{tr}(\boldsymbol{\beta}\boldsymbol{\beta}'\Lambda) - \text{tr} \{ \boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{S}(\mathbf{t}) \}. \tag{6}$$

2.4. A joint design for functional data

In practice, there might be multiple goals in design with each goal resulting in one optimal design. Depending on the goal, the corresponding optimal design may vary and may not be optimal for alternative goals. Indeed, the optimal sampling points for predicting functions may not be the optimal sampling points for predicting a scalar outcome, and vice versa. It may thus be useful to consider a joint design to balance between the different goals. Note that joint designs may also be referred to as compound designs in the statistical design literature (Atkinson et al., 2007, Chapter 21).

Before formulating a joint design, consider first the design objective function

$$\mathcal{M}_{\mathbf{B}}(\mathbf{t}) = \text{tr}(\mathbf{B}\mathbf{A}) - \text{tr}\{\mathbf{B}\mathbf{S}(\mathbf{t})\}, \quad (7)$$

where \mathbf{B} is an arbitrary positive semidefinite matrix and we will call it a “linear design criterion matrix” as the objective function depends linearly on elements of \mathbf{B} . The form in (7) is general with different \mathbf{B} leading to different designs. In particular, it includes as special cases, the objective functions for the design for predicting functions (5) and for the design for predicting an outcome (6). Indeed, for predicting the growth curve of a new subject, \mathbf{B} is the identity matrix and for predicting a scalar outcome of a new subject, $\mathbf{B} = \boldsymbol{\beta}\boldsymbol{\beta}'$. Additionally, if it is more important to predict a curve more accurately at some time points than others, one may consider a weighted mean integrated squared error $E \int_{\mathcal{T}} w(t)[X_{i^*}(t) - E\{X_{i^*}(t)|\mathbf{W}_{i^*}(\mathbf{t})\}]^2 dt$ for some known weight function $0 \leq w(t) \leq 1$. It can be shown that the objective function to be minimized still takes the form in (7) with a particular design criterion matrix. Specifically, $[\mathbf{B}]_{\ell, \ell'} = \int w(t)\phi_{\ell}(t)\phi_{\ell'}(t)dt$ and is positive semidefinite with a finite operator norm by Lemma 1 in Appendix A.

Now consider a bivariate continuous function $f(\cdot, \cdot)$ on $[0, \infty) \times [0, \infty)$ and the objective function $\mathcal{M}(\mathbf{t}) = f(\mathcal{M}_1(\mathbf{t}), \mathcal{M}_2(\mathbf{t}))$. Suppose the joint design is to minimize the objective function $\mathcal{M}(\mathbf{t})$. It is reasonable to impose the following assumption on $f(\cdot, \cdot)$:

Assumption 1. $f(x, y)$ is nondecreasing along both x and y and $f(0, 0) = 0$. Moreover, $\lim_{x \rightarrow 0, y \rightarrow 0} f(x, y) = 0$.

Let w_1 and w_2 be two fixed non-negative constants. Two sensible forms of f are: $f_1(x, y) = w_1x + w_2y$ and $f_2(x, y) = \max(w_1x, w_2y)$. The former is a joint design that minimizes a linear combination of two prediction errors while the latter means that the joint design aims to minimize the maximum of the two prediction errors (up to multiplicative weights). In particular, $f_1(\mathcal{M}_1(\mathbf{t}), \mathcal{M}_2(\mathbf{t})) = \mathcal{M}_{\mathbf{B}}(\mathbf{t})$ with $\mathbf{B} = w_1\mathbf{I} + w_2\boldsymbol{\beta}\boldsymbol{\beta}'$. It is straightforward to show that both forms satisfy Assumption 1. The two constants w_1 and w_2 are used to control the weights of the two different design objective functions. One reasonable choice of w_1 and w_2 is to balance the two design objective functions such that one design does not dominate the other. In view of (7) and Theorem 2 from the following section, we may let $w_1 = 1/\text{tr}(\mathbf{A})$ and $w_2 = 1/\text{tr}(\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{A})$, and it can then be shown that $0 \leq w_1\mathcal{M}_1(\mathbf{t}) \leq 1$ and $0 \leq w_2\mathcal{M}_2(\mathbf{t}) \leq 1$.

3. Properties of $\mathcal{M}(\mathbf{t})$

In this section we study the properties of $\mathcal{M}(\mathbf{t})$ for any function $f(\cdot, \cdot)$ that satisfies Assumption 1. We assume that the random functions, $X(t)$, are square integrable (i.e., $E\{\int_{\mathcal{T}} X(t)^2 dt\} < \infty$) and the coefficient function $\beta(t)$ in the functional linear regression model (2) is also square integrable (i.e., $\int_{\mathcal{T}} \beta^2(t) dt < \infty$). Proofs of the theorems are provided in Appendix A.

Theorem 1. Suppose $\mathbf{t} \in [0, 1]^p$, $\tilde{\mathbf{t}} \in [0, 1]^{p+c}$ for some fixed integer $c > 0$ and $\mathbf{t} \subset \tilde{\mathbf{t}}$, then $\mathcal{M}(\tilde{\mathbf{t}}) \leq \mathcal{M}(\mathbf{t})$.

Theorem 1 implies that more observations (i.e., larger p) do not increase the value of the objective function $\mathcal{M}(\mathbf{t})$. We also study the deterministic bound of $\mathcal{M}(\mathbf{t})$ as p diverges to infinity according to a fixed design.

Theorem 2. Suppose that the assumptions stated in Appendix A hold. For the fixed design where $\mathbf{t}_p = \{0, 1/p, 2/p, \dots, (p-1)/p\}'$, we have $\lim_{p \rightarrow \infty} \mathcal{M}(\mathbf{t}_p) = 0$.

Theorem 2 provides the rationale that a dense set of time points in \mathcal{T} is sufficient as the candidate sampling points. In practice, because of the cost for data collection and other considerations, a small number of sampling points with reasonable prediction power might be preferred. In Section 4.2, we propose a data-driven method for selecting the number of optimal time points.

4. Implementation

4.1. Model estimation using pilot data

To implement the proposed optimal design, we need to estimate the covariance function $r(s, t)$, error variance σ_{ϵ}^2 and coefficient function $\beta(t)$ using pilot data. Many methods exist for covariance function estimation including local polynomial regression (Yao et al., 2005), mixed effects models (James et al., 2000) and geometric PCA (Peng and Paul, 2009). We use the fast covariance estimation method (FACES) from Xiao et al. (2017), which uses a penalized tensor product of cubic B-splines for approximating the true covariance function. The error variance σ_{ϵ}^2 can also be estimated by FACES. As for estimating $\beta(t)$, we select K , the number of eigenfunctions, by the percentage of variance explained (PVE) with a value of 0.95.

4.2. Optimization algorithm and selection of number of design points

In practice, the optimal sampling points are selected from a pre-determined set of candidate time points, denoted by \mathbf{s} . [Theorem 2](#) suggests that equally spaced sampling points can form a reasonable set of candidate points. If the number of selected design points is small, then we use a full search algorithm (i.e., we evaluate \mathcal{M} for every combination of p points from \mathbf{s}). If the number of selected design points is large, a full search becomes computationally difficult and one may use a Monte Carlo sampling method in [Wu et al. \(2017\)](#) or a sequential search method in [Ji and Müller \(2017\)](#). In this paper we focus on the full search algorithm.

In many applications, the number of optimal time points p may not be known *a priori*. One approach is to choose the smallest p such that the expected error is smaller than some pre-determined tolerance error. Alternatively, similar to [Ferraty et al. \(2010\)](#), one may incorporate the cost of collecting more sampling points into consideration. Here we propose a new method for selecting p . First, when $\mathbf{t} = \emptyset$, an empty set, we define $\mathcal{M}_1(\emptyset) = \text{tr}(\Lambda)$, $\mathcal{M}_2(\emptyset) = \text{tr}(\beta\beta'\Lambda)$, and $\mathcal{M}(\emptyset) = f(\mathcal{M}_1(\emptyset), \mathcal{M}_2(\emptyset))$. Note that $\mathcal{M}_1(\emptyset)$ is the total variation of the functional predictor while $\mathcal{M}_2(\emptyset)$ is the total variation of the response that can be explained by the functional predictor. Then it can be easily verified by the definitions in [\(5\)](#) and [\(6\)](#) and the assumptions on f that $\mathcal{M}(\mathbf{t}) \leq \mathcal{M}(\emptyset)$ for any \mathbf{t} of any dimension. Let $\mathbf{t}_p^* = \arg \min_{\mathbf{t} \subseteq \mathbf{s}, \mathbf{t} \in \mathcal{T}^p} \mathcal{M}(\mathbf{t})$. Then,

$$p^* := \min_{p \in \mathbb{N}} \{ \mathcal{M}(\mathbf{t}_p^*) / \mathcal{M}(\emptyset) + \delta p \}, \quad (8)$$

where $0 < \delta < 1$ is a fixed constant corresponding to the maximum percent reduction in expected squared error gained by augmenting the design with an additional design point.

By [Theorem 1](#) in [Section 3](#), for any p , $\mathcal{M}(\mathbf{t}_{p+1}^*) \leq \mathcal{M}(\mathbf{t}_p^*) \leq \mathcal{M}(\emptyset)$. Thus, the relative error level $\mathcal{M}(\mathbf{t}_p^*) / \mathcal{M}(\emptyset)$ is a decreasing function of p and converges to 0 by [Theorem 2](#). This implies that for any fixed $\delta > 0$, p^* is finite. In practice, we plug estimated model components (see [Section 4.1](#)) into \mathcal{M} to obtain $\widehat{\mathcal{M}}$. Then we let $\widehat{\mathbf{t}}_p^* = \arg \min_{\mathbf{t} \subseteq \mathbf{s}, \mathbf{t} \in \mathbb{R}^p} \widehat{\mathcal{M}}(\mathbf{t})$ and define

$$\widehat{p}^* := \min_{p \in \mathbb{N}} \{ \widehat{\mathcal{M}}(\widehat{\mathbf{t}}_p^*) / \widehat{\mathcal{M}}(\emptyset) + \delta p \}. \quad (9)$$

Small values of δ seem preferable and we use $\delta = 0.05$ in both the data application and simulations. This implies that we select a \widehat{p} such that the addition of a new design point will result in no more than 5% of reduction in expected squared error with respect to the error reduced by using the fully observed functional predictor.

4.3. Software and shiny interactive graphic

The proposed optimal design method has been implemented as an R package ([R Core Team, 2016](#)) `FDAdesign` that includes interactive graphics using `shiny` ([Chang et al., 2016](#)) which can be used to evaluate design objectives corresponding to different sampling designs. The interface of the graphic is illustrated in the data application. Details about using the `FDAdesign` package and the interactive graphics can be found in [Section S.1](#) of the Supplementary materials.

5. Application to fetal ultrasound

We apply the proposed methodology to fetal growth data, where ultrasound scans were performed at different weeks of gestational age (GA). For this analysis, we model measurements of abdominal circumference by ultrasound (scaled within the range of 0 and 1) to estimate individual fetal growth trajectories and use newborn birth weight as the scalar outcome.

The fetal growth dataset contains between 1 and 6 ultrasound scans for each of 2388 subjects, with most subjects having 5 scans. The spaghetti plot for the ultrasound measurements is shown in [Fig. 1](#), with data from 3 subjects highlighted. While the 3 subjects do show some degree of curvilinearity, the overall pattern of trajectories raises the question of whether a linear mixed effects model would suffice for this data.

Thus, we compare the functional model and the linear mixed effects model using 10-fold cross validation. We find that the linear mixed effects model has twice the prediction error of the functional model. [Figure S.2](#) in the Supplementary materials illustrates the prediction performance of the two models for one particular case, where 90% data are used for model estimation and the remaining 10% data are used for evaluation. Therefore, using functional model seems more appropriate for this application.

[Fig. 2](#) displays the fPCA fit. The top panels of [Fig. 2](#) show that both the estimated mean function and variance function are increasing with gestational age. The bottom left panel of [Fig. 2](#) indicates high positive correlations (>0.8) when both gestational ages are smaller than 32 weeks. The top three estimated eigenvalues are 8.7×10^{-4} , 1.0×10^{-4} and 0.3×10^{-5} , respectively, with the corresponding estimated eigenfunctions shown in the bottom right panel of [Fig. 2](#). The estimated error variance is 2.2×10^{-4} .

When we predict subject birth weight using the functional linear model [\(2\)](#) with abdominal circumference as the functional covariate, it turns out that about 99% of the variation in the birth weight is explained by the functional covariate. For the estimated coefficient function; see [Figure S.3](#) of the Supplementary materials.

Finally we consider a linear joint design with the target of accurately recovering the ultrasound measurements of fetal abdominal circumference and predicting the newborn outcome of birth weight. The objective function for the joint design

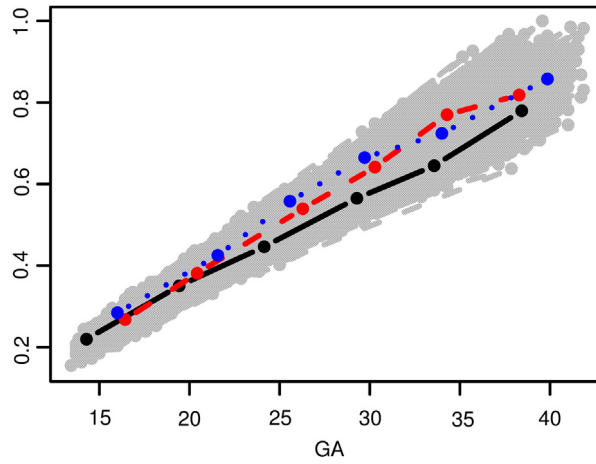


Fig. 1. Spaghetti plot of the fetal ultrasound data.

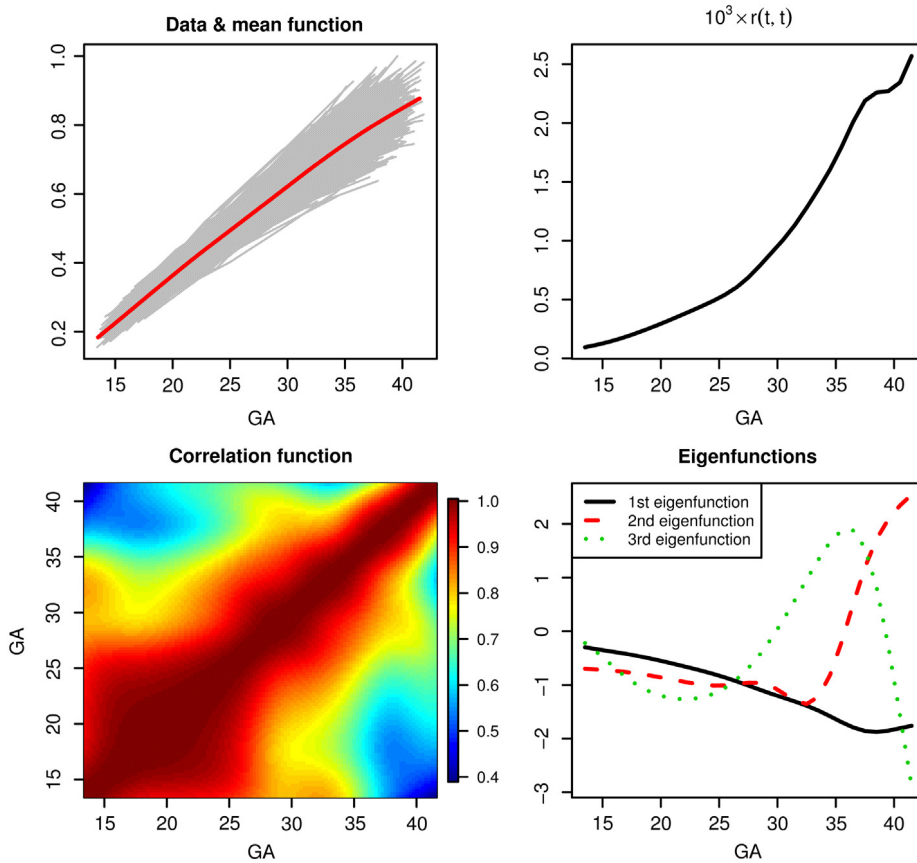


Fig. 2. fPCA fit to the ultrasound measurements.

is $w_1 \hat{\mathcal{M}}_1(\mathbf{t}) + w_2 \hat{\mathcal{M}}_2(\mathbf{t})$, where $\hat{\mathcal{M}}_1(\mathbf{t})$ is the estimated objective function for recovering individual functions while $\hat{\mathcal{M}}_2(\mathbf{t})$ is the estimated objective function for predicting a scalar outcome; see Section 2.4 for more details. To balance the two objective functions, we let $w_1 = (\sum_{\ell=1}^3 \hat{\lambda}_\ell)^{-1}$ and $w_2 = (\sum_{\ell=1}^3 \hat{\lambda}_\ell \hat{\beta}_\ell^2)^{-1}$, where the $\hat{\lambda}_\ell$ and the $\hat{\beta}_\ell$ are estimated from the fetal data and the top 3 eigenvalues are selected using a PVE of 0.95. The above weights ensure that $0 \leq w_1 \hat{\mathcal{M}}_1(\mathbf{t}) \leq 1$ and $0 \leq w_2 \hat{\mathcal{M}}_2(\mathbf{t}) \leq 1$. As a result, one objective function will not dominate the other.

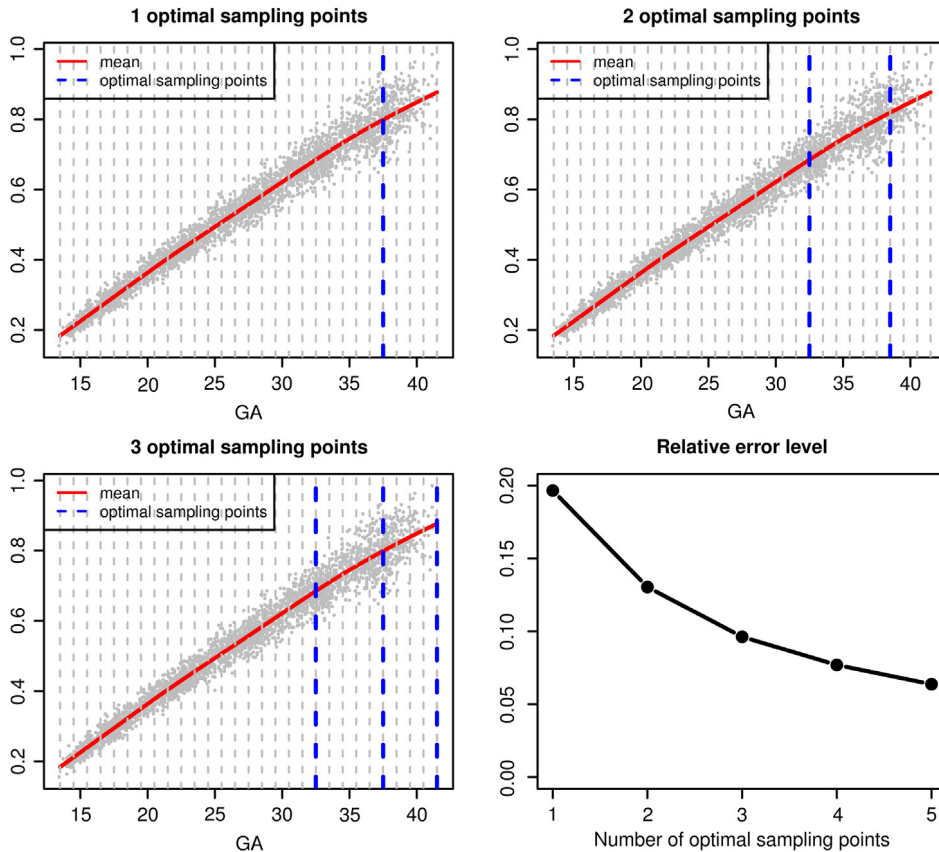


Fig. 3. Optimal sampling points and the corresponding relative error levels. The dashed gray vertical lines in the top left, top right and bottom left panels are the candidate sampling points. The relative error levels are $\widehat{\mathcal{M}}(\hat{\mathbf{t}}_p)/\widehat{\mathcal{M}}(\emptyset)$, where $\hat{\mathbf{t}}_p$ is the vector of p optimal sampling points determined by $\widehat{\mathcal{M}}$.

We let the set of candidate time points \mathbf{s} be the collection of half weeks between 13 and 41 weeks gestational age. Using the proposed method, we determine the optimal sampling points when the number of sampling points p is fixed at 1, 2 and 3. We also calculate the relative error $\widehat{\mathcal{M}}/\widehat{\mathcal{M}}(\emptyset)$ and Fig. 3 displays the results. The top left panel of Fig. 3 shows that if only 1 sampling point is selected, then 37 weeks is the optimal time point for collecting the ultrasound measurement and its relative error is about 0.20 (bottom right panel of Fig. 3). If 2 sampling points are desired, then 32 and 38 weeks are the optimal time points for collecting ultrasound measurements. With 2 optimal sampling points, the relative error is 0.13, which is smaller than the relative error with only 1 optimal sampling point. The bottom right panel of Fig. 3 displays the relative error with several values of p . As expected, the relative error decreases as p increases. Using the selection criterion (9) with $\delta = 0.05$, we determine that optimally, 2 sampling points would be selected.

To evaluate the uncertainty in the estimated optimal sampling points, we bootstrap the fetal ultrasound data at the subject level and select the optimal sampling points for 1000 bootstrapped datasets. Fig. 4 gives the histograms of the selected optimal sampling points, which show small variability of the estimated optimal sampling points. For example, for $p = 1$, week 37 is selected about 60% of the times.

Finally, we plot in Fig. 5 screenshots of the Shiny interface for the fetal ultrasound. The top panel displays the heat map of the objective function/prediction errors as a bivariate function of two scan weeks. The optimal weeks are highlighted. The heat map indicates that at least one sampling point needs to be no early than 33 weeks in order to obtain a relatively small prediction error. As these plots evaluate the prediction error of any combination of candidate sampling points, they can be used to find all candidate sampling points that give a prediction error smaller than certain fixed error. The interface is interactive as users can select the first scan weeks and then the application will find the optimal second scan weeks. Moreover, users can go further by selecting the second scan weeks and compare the results with the optimal scan weeks. For example, as illustrated in the bottom panel, 13 weeks is selected for the first scan, then the 37 weeks is found to be optimal second scan weeks (left plot in the bottom panel). If 16 weeks is also selected for the second scan, then the result can be compared with several different choices including the optimal scan weeks (right plot in the bottom panel). A similar screenshot with the goal of selecting just one scan is presented in Figure S.4 of the Supplementary materials.

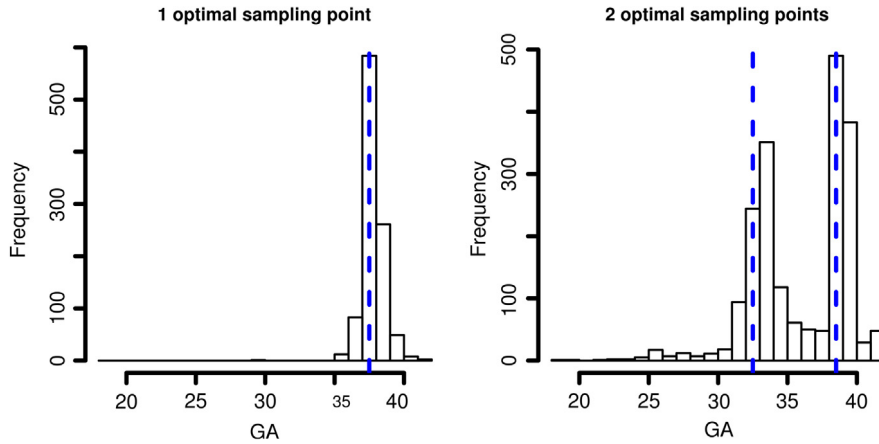


Fig. 4. Histograms of selected optimal scan weeks from 1000 bootstrapped datasets for $p = 1$ and $p = 2$. The blue dashed lines are the estimated optimal scan weeks using the original fetal ultrasound data.

6. A simulation study

We conduct a simulation study to investigate the performance of the proposed design for (a) estimating optimal sampling points and for (b) selecting the number of optimal sampling points. We also compare functional data models against a parametric mixed effects model in terms of estimating optimal sampling points, when data is generated from either a functional data model or a parametric mixed effects model. We focus on the linear joint design where the goal is to best predict both an underlying true curve and a scalar outcome and we use the same design criterion matrix in the data example with weights $w_1 = (\sum_{\ell} \lambda_{\ell})^{-1}$ and $w_2 = (\sum_{\ell} \lambda_{\ell} \beta_{\ell}^2)^{-1}$.

6.1. Simulation settings

For each simulation scenario, we use 200 Monte Carlo samples from the model in (1). For simplicity, we let the mean function $\mu(t)$ be zero for all t . We generate $X_i(t)$ by $X_i(t) = \sum_{\ell=1}^5 \xi_{i\ell} \phi_{\ell}(t)$, where $\{\phi_1(t), \dots, \phi_5(t)\}$ is a set of orthonormal eigenfunctions (to be specified later) and $\xi_{i\ell}$ is sampled from a normal distribution with mean zero and variance $\lambda_{\ell} = 10/2^{\ell}$. Random errors ϵ_{ij} are sampled independently from a normal distribution with mean zero and variance $\sigma_{\epsilon}^2 = 9.6875$, which implies that the signal to noise ratio $\sigma_{\epsilon}^{-2} \sum_{\ell=1}^5 \lambda_{\ell}$ equals one. The number of observations per subject varies across subjects and the sampling time points are drawn from the uniform distribution in the unit interval.

We consider a factorial design with three experimental factors:

- E1. Covariance function $r(s, t)$:
 - (a) *Periodic* covariance $r(s, t)$ induced by five Fourier bases: $\phi_{\ell}(t) = \sqrt{2} \sin((\ell + 1)\pi t)$ for odd ℓ and $\phi_{\ell}(t) = \sqrt{2} \cos(\ell\pi t)$ for even ℓ . The covariance function is periodic because $r(s, t) = r(1 - s, t)$.
 - (b) *Non-periodic* covariance $r(s, t)$ induced by five eigenfunctions shown in Figure S.6 of the Supplementary materials and the eigenfunctions do not have analytical forms.
- E2. Number of observations per subject:
 - (a) $m_i \sim \text{Uniform}\{3, 4, 5\}$ and (b) $m_i \sim \text{Uniform}\{7, \dots, 10\}$.
- E3. Number of subjects: (a) $n = 400$; (b) $n = 800$; and (b) $n = 1500$.

Thus, in total there are 12 model conditions to examine.

Then the scalar outcomes, Y , are generated from the functional linear model in (2). For simplicity, we let intercept $\alpha = 0$. We use four different coefficient functions $\beta(t)$ (see Figure S.7 of the Supplementary materials):

FLM-Case1 $\beta(t) = \sum_{\ell=1}^5 \beta_{\ell} \phi_{\ell}(t)$, with $\beta_{\ell} = 4, 2.5, 1.5, 1$, and 0.5 .

FLM-Case2 $\beta(t) = 12t^2$.

FLM-Case3 $\beta(t) = 4(t - 0.5)^2 + 4\sin(2\pi t) + 2\cos(6\pi t)$.

Note that in FLM-Case1 the coefficient function $\beta(t)$ depends on the eigenfunctions $\phi_{\ell}(t)$ and is different for the periodic and non-periodic covariances (see Figure S.7 of the Supplementary materials). Random errors in (2) were sampled independently from a normal distribution with mean zero and variance $\sigma_{\epsilon}^2 = 4$.

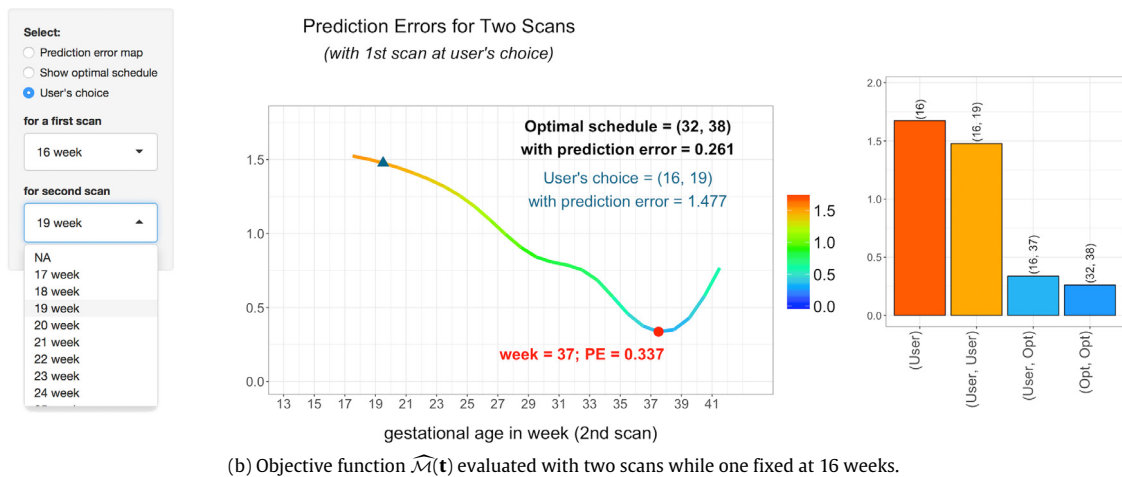
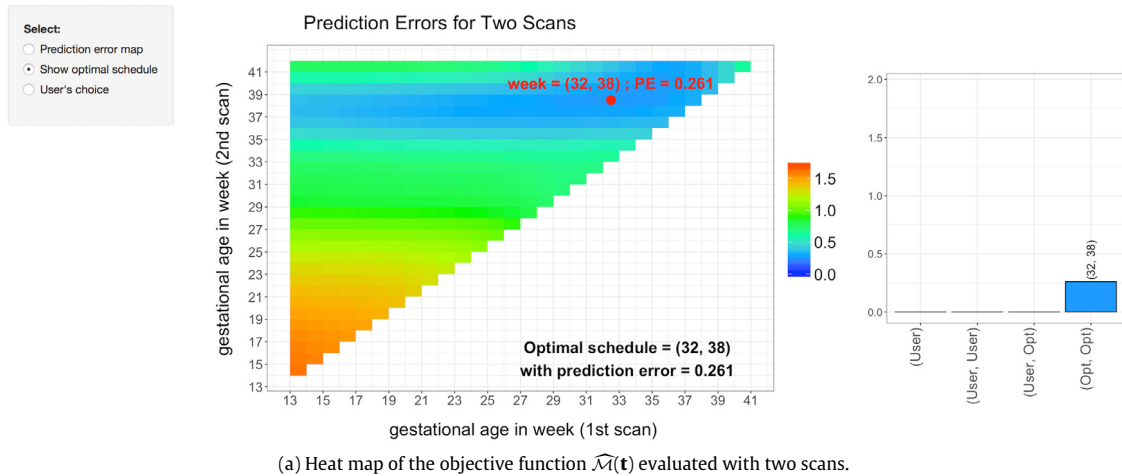


Fig. 5. Screenshots of the interface of the Shiny application for the fetal ultrasound.

6.2. Results for estimation of optimal sampling points

We consider estimation of optimal sampling points when the number of optimal points p is fixed at either 3, 4 or 5. Let \mathbf{t}_p^* be the p optimal sampling points that minimize the true objective function $\mathcal{M}(\cdot)$ and let $\widehat{\mathbf{t}}_p$ be the p selected sampling points that minimize the estimated objective function $\widehat{\mathcal{M}}(\cdot)$. We evaluate the accuracy of the estimated optimal sampling points using the following evaluation criterion:

$$ARE_{p, i_{sim}} = \frac{|\mathcal{M}(\mathbf{t}_p^*) - \mathcal{M}(\widehat{\mathbf{t}}_{p, i_{sim}})|}{\mathcal{M}(\mathbf{t}_p^*)}. \tag{10}$$

The absolute relative error, $ARE_{p, i_{sim}}$, measures how close the expected (integrated) squared error using observations collected at the p estimated optimal sampling points is to the expected (integrated) squared error using the p true optimal points. We compare between $\mathcal{M}(\mathbf{t}_p^*)$ and $\mathcal{M}(\widehat{\mathbf{t}}_{p, i_{sim}})$, rather than between $\widehat{\mathbf{t}}_p$ and \mathbf{t}_p^* , for the following reasons. First, when the covariance function $r(s, t)$ is periodic as the one shown in the top left panel of Figure S.6 of the Supplementary materials, \mathbf{t}_p^* is not identifiable. This is because with a periodic covariance function, data (excluding random errors) collected at any sampling point in the left half of the domain is the same as data collected at one sampling point in the right half. The identifiability issue is illustrated in Section S.3.2 of the Supplementary materials. Second, as our ultimate goal is to minimize $\mathcal{M}(\cdot)$, the expected (integrated) squared error, we consider that the measure $ARE_{p, i_{sim}}$ is more appropriate.

In addition to functional data methods, we consider the following linear mixed effects (LME) model, $W_{ij} = b_{i0} + b_{i1}t_{ij} + \epsilon_{ij}$, for estimating the covariance function $r(s, t)$, where b_{i0} and b_{i1} are subject-specific random intercept and slope, respectively. The above model leads to a quadratic covariance function. In the following tables we use the labels, *non-parametric* and

Table 1

Median of absolute relative errors, $\{ARE_{p,i_{sim}} : i_{sim} = 1, \dots, 200\}$ and the corresponding interquartile ranges (IQR) in parentheses for the case of the periodic covariance.

| | | Joint-Case1 | | Joint-Case2 | | Joint-Case3 | |
|-----------------|-------------------------------------|----------------|---------------|----------------|---------------|----------------|---------------|
| | | Non-parametric | Parametric | Non-parametric | Parametric | Non-parametric | Parametric |
| <i>p</i> = 3 | | | | | | | |
| <i>n</i> = 400 | <i>m_i</i> ~ {3, 4, 5} | 0.019 (0.019) | 1.292 (0.032) | 0.018 (0.033) | 1.507 (0.284) | 0.055 (0.053) | 1.598 (0.006) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.011 (0.016) | 1.537 (0.264) | 0.010 (0.015) | 1.507 (0.284) | 0.024 (0.026) | 1.598 (0.006) |
| <i>n</i> = 800 | <i>m_i</i> ~ {3, 4, 5} | 0.011 (0.016) | 1.537 (0.000) | 0.010 (0.023) | 1.507 (0.000) | 0.030 (0.037) | 1.598 (0.006) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.005 (0.011) | 1.537 (0.264) | 0.009 (0.007) | 1.507 (0.284) | 0.012 (0.014) | 1.598 (0.006) |
| <i>n</i> = 1500 | <i>m_i</i> ~ {3, 4, 5} | 0.011 (0.014) | 1.537 (0.000) | 0.010 (0.014) | 1.507 (0.000) | 0.026 (0.024) | 1.601 (0.006) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.005 (0.007) | 1.537 (0.264) | 0.005 (0.007) | 1.507 (0.284) | 0.012 (0.010) | 1.598 (0.006) |
| <i>p</i> = 4 | | | | | | | |
| <i>n</i> = 400 | <i>m_i</i> ~ {3, 4, 5} | 0.047 (0.030) | 1.585 (0.175) | 0.046 (0.037) | 1.612 (0.286) | 0.070 (0.054) | 1.983 (0.099) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.015 (0.023) | 1.676 (0.000) | 0.016 (0.023) | 1.612 (0.000) | 0.027 (0.035) | 1.983 (0.000) |
| <i>n</i> = 800 | <i>m_i</i> ~ {3, 4, 5} | 0.031 (0.040) | 1.676 (0.264) | 0.029 (0.043) | 1.612 (0.286) | 0.045 (0.041) | 1.983 (0.025) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.012 (0.013) | 1.676 (0.000) | 0.011 (0.011) | 1.612 (0.000) | 0.02 (0.022) | 1.983 (0.000) |
| <i>n</i> = 1500 | <i>m_i</i> ~ {3, 4, 5} | 0.018 (0.024) | 1.676 (0.000) | 0.018 (0.026) | 1.612 (0.000) | 0.033 (0.032) | 1.983 (0.000) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.007 (0.010) | 1.676 (0.000) | 0.007 (0.010) | 1.612 (0.000) | 0.008 (0.016) | 1.983 (0.000) |
| <i>p</i> = 5 | | | | | | | |
| <i>n</i> = 400 | <i>m_i</i> ~ {3, 4, 5} | 0.059 (0.051) | 1.713 (0.050) | 0.051 (0.044) | 1.929 (0.342) | 0.063 (0.070) | 2.167 (0.020) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.027 (0.027) | 1.695 (0.321) | 0.022 (0.027) | 1.587 (0.342) | 0.026 (0.028) | 2.167 (0.020) |
| <i>n</i> = 800 | <i>m_i</i> ~ {3, 4, 5} | 0.039 (0.037) | 2.016 (0.321) | 0.039 (0.037) | 1.929 (0.342) | 0.043 (0.041) | 2.167 (0.020) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.020 (0.022) | 1.695 (0.321) | 0.013 (0.020) | 1.587 (0.342) | 0.016 (0.016) | 2.167 (0.020) |
| <i>n</i> = 1500 | <i>m_i</i> ~ {3, 4, 5} | 0.029 (0.027) | 2.016 (0.321) | 0.029 (0.031) | 1.929 (0.342) | 0.032 (0.036) | 2.167 (0.020) |
| | <i>m_i</i> ~ {7, ..., 10} | 0.009 (0.017) | 1.695 (0.000) | 0.007 (0.009) | 1.587 (0.000) | 0.010 (0.015) | 2.167 (0.020) |

Note: **Joint-Case1** indicates that the scalar responses are generated using $\beta(t)$ in FLM-Case1; similarly, **Joint-Case2** corresponds to FLM-Case2 and **Joint-Case3** to FLM-Case3. *non-parametric* and *Parametric* refer to the covariance estimation using the fPCA and LME models, respectively.

parametric, to indicate covariance estimation using the functional data model and using the linear mixed effects model, respectively.

The results with the periodic covariance function are summarized in Table 1. The proposed design works well and the ARE decreases as a function of number of subjects *n* and number of observations per subject *m_i*. The improved performance is due to improved estimation accuracy of the covariance function (and associated eigenfunctions and eigenvalues) as well as of the error variance of the random errors (results not shown). The results with the non-periodic covariance function are similar and are shown in Section S.3.3 of the Supplementary materials.

In addition to the ARE measure, we study the behavior of the objective function \mathcal{M} for different choices of *p* by investigating the median and interquartile range (IQR) of the $\mathcal{M}(\hat{\mathbf{t}}_{p,i_{sim}})$. The statistics for the periodic and non-periodic covariance cases are presented in Tables S.1 and S.2 of the Supplementary materials, respectively. The true objective function $\mathcal{M}(\cdot)$ depends on the true covariance function $r(s, t)$, *n*, *m_i*, and $\beta(t)$. Thus $\mathcal{M}(\hat{\mathbf{t}}_{p,i_{sim}})$ can only be compared across different *p*, but not across different simulation settings with different *n* or *m_i*.

As expected $\mathcal{M}(\hat{\mathbf{t}}_p)$ decreases with more number of optimal sampling points. The same holds when we use the *parametric* covariance estimation, estimating $r(s, t)$ using the LME model. Because the LME model is misspecified for modeling functional data, selecting more optimal points by the *parametric* estimation has only a slight effect on improving the prediction accuracy. In all cases the proposed method with the *non-parametric* covariance estimation gives a smaller prediction error than the *parametric* estimation. When data are generated from the LME model, the proposed method performs equally well with both the *non-parametric* and *parametric* covariance estimation; see Section S.3.4 of the Supplementary materials. In conclusion, the proposed method with *non-parametric* covariance estimation using the fPCA model performs well on data with both simple and complex covariance structures.

6.3. Results for selection of number of optimal sampling points

Now we evaluate the performance of the proposed method in (9) for selecting the number of optimal sampling points *p*. We use $\delta = 0.05$ and the true number of optimal points p^* determined by (8) is 3. The performance of the proposed method is assessed in terms of the proportion of selecting the correct number of optimal sampling points, $N_{sim}^{-1} \sum_{i_{sim}=1}^{N_{sim}} 1_{\{\hat{p}_{i_{sim}}^* = 3\}}$, where $1_{\{\cdot\}}$ is an indicator function and $\hat{p}_{i_{sim}}^*$ is the number of optimal sampling points determined by (9) using the *i*th simulated data.

The simulation results are presented in Table 2. We see that the performance of the proposed method is excellent for all cases. The results for the non-periodic function are similarly good and presented in Table S.4 of the Supplementary materials.

Table 2

Proportion of selected number of points being equal to 3 for the case of the periodic covariance.

| | | Joint-Case1 | Joint-Case2 | Joint-Case3 |
|----------|-----------------------------|-------------|-------------|-------------|
| n = 400 | $m_i \sim \{3, 4, 5\}$ | 0.94 | 0.97 | 0.94 |
| | $m_i \sim \{7, \dots, 10\}$ | 0.98 | 0.99 | 0.95 |
| n = 800 | $m_i \sim \{3, 4, 5\}$ | 0.98 | 0.99 | 0.95 |
| | $m_i \sim \{7, \dots, 10\}$ | 0.99 | 1.00 | 0.98 |
| n = 1500 | $m_i \sim \{3, 4, 5\}$ | 0.99 | 0.99 | 0.96 |
| | $m_i \sim \{7, \dots, 10\}$ | 1.00 | 1.00 | 0.99 |

Note: **Joint-Case1** indicates that the scalar responses are generated using $\beta(t)$ in FLM-Case1; similarly, **Joint-Case2** corresponds to FLM-Case2 and **Joint-Case3** to FLM-Case3.

6.4. Uncertainty of estimated optimal sampling points

To assess the uncertainty of the optimal sampling points estimated from the proposed method, we use a bootstrap approach as in the data application. For each simulated data, we bootstrap at the subject level, select optimal sampling points from the model estimation based on the bootstrapped data, and calculate the third quartile and 90% percentile of absolute relative errors in (10). The medians of the percentiles are presented in Tables S.7 and S.8 of the Supplementary materials. The results show good stability of the estimated optimal sampling points, which gets better when either the sample size or the number of observations per subject increases.

Acknowledgment

This work was supported by Grant Number OPP1148351 from the Bill & Melinda Gates Foundation. This work represents the opinions of the researchers and not necessarily that of the granting organizations.

Appendix A. Assumptions and proofs

Assumptions for Theorem 2. Similar to Bunea and Xiao (2015), we make the following assumptions on the covariance function $r(s, t)$:

Assumption A. $r(s, t)$ is continuous and a positive semi-definite kernel.

Assumption B. $\sup_k \sup_{s \in \mathcal{T}} |\phi_k(s)|$ is bounded by a constant.

Assumption C. $\frac{\partial r(t,t)}{\partial t}$ is continuous in \mathcal{T} . Moreover, $\int |\frac{\partial r(t,t)}{\partial t}| dt$ is finite.

Assumption D. For all k , $\sup_{s \in \mathcal{T}} |\phi_k^{(1)}(s)| \leq Ck^{\gamma_1}$, where $\phi_k^{(1)}(s)$ is the derivative of ϕ_k and C, γ_1 are positive constants.

Assumption E. There exist constants C_1, C_2 and $\beta_2 \geq \beta_1 > \max(\gamma_1, 1)$ such that $C_2k^{-\beta_2} \leq \lambda_k \leq C_1k^{-\beta_1}$, for all k . Moreover, there exist constants C_3 and $\beta_3 > \beta_2$ such that $\lambda_k - \lambda_{k+1} \geq C_3k^{-\beta_3}$, for all k .

Finally, we assume that the linear design matrix \mathbf{B} is positive semidefinite with a finite operator norm.

Lemma 1. If $0 \leq w(t) \leq 1$ for $t \in \mathcal{T}$ for a bounded interval \mathcal{T} . Then $\mathbf{B} = [b_{\ell, \ell'}]_{1 \leq \ell, \ell' \leq \infty}$ with $b_{\ell, \ell'} = \int w(t)\phi_\ell(t)\phi_{\ell'}(t)dt$ is positive semidefinite with a finite operator norm.

Proof. It suffices to prove that $\mathbf{x}'\mathbf{B}\mathbf{x}$ is non-negative and finite for any $\mathbf{x} = (x_1, x_2, \dots, x_K)'$ with $\sum_\ell x_\ell^2 = 1$. Indeed,

$$\begin{aligned} \mathbf{x}'\mathbf{B}\mathbf{x} &= \sum_\ell \sum_{\ell'} x_\ell x_{\ell'} \int w(t)\phi_\ell(t)\phi_{\ell'}(t)dt = \int w(t) \sum_\ell \sum_{\ell'} x_\ell x_{\ell'} \phi_\ell(t)\phi_{\ell'}(t)dt \\ &= \int w(t) \left\{ \sum_\ell x_\ell \phi_\ell(t) \right\}^2 dt \geq 0, \end{aligned}$$

and furthermore

$$\begin{aligned} \mathbf{x}'\mathbf{B}\mathbf{x} &= \int w(t) \left\{ \sum_\ell x_\ell \phi_\ell(t) \right\}^2 dt \leq \int \left\{ \sum_\ell x_\ell \phi_\ell(t) \right\}^2 dt \\ &\leq \sum_\ell \sum_{\ell'} x_\ell x_{\ell'} \int \phi_\ell(t)\phi_{\ell'}(t)dt \leq \sum_\ell x_\ell^2 = 1. \quad \square \end{aligned}$$

Proof of Theorem 1. In light of Assumption 1 and Eq. (7), it suffices to prove the theorem for the design objective function in (7) with an arbitrary positive semidefinite matrix \mathbf{B} . Define the symmetric matrix $\mathbf{F}(\mathbf{t})$ as in Lemma 2, then $\mathbf{F}(\hat{\mathbf{t}}) - \mathbf{F}(\mathbf{t})$ is

positive semidefinite and $\mathbf{S}(\mathbf{t}) = \Lambda^{1/2}\mathbf{F}(\mathbf{t})\Lambda^{1/2}$ and $\mathbf{S}(\tilde{\mathbf{t}}) = \Lambda^{1/2}\mathbf{F}(\tilde{\mathbf{t}})\Lambda^{1/2}$. For simplicity we denote $\mathbf{F}(\mathbf{t})$ by \mathbf{F} and $\mathbf{F}(\tilde{\mathbf{t}})$ by $\tilde{\mathbf{F}}$. Since $\tilde{\mathbf{F}} - \mathbf{F}$ is positive semidefinite, there exists a matrix $(\tilde{\mathbf{F}} - \mathbf{F})^{1/2}$ such that $\tilde{\mathbf{F}} - \mathbf{F} = (\tilde{\mathbf{F}} - \mathbf{F})^{1/2}(\tilde{\mathbf{F}} - \mathbf{F})^{1/2}$. Then

$$\begin{aligned} \mathcal{M}(\tilde{\mathbf{t}}) - \mathcal{M}(\mathbf{t}) &= \text{tr}(\mathbf{B}\Lambda^{1/2}\tilde{\mathbf{F}}\Lambda^{1/2} - \mathbf{B}\Lambda^{1/2}\mathbf{F}\Lambda^{1/2}) \\ &= \text{tr}\{\mathbf{B}\Lambda^{1/2}(\tilde{\mathbf{F}} - \mathbf{F})\Lambda^{1/2}\} \\ &= \left\| (\tilde{\mathbf{F}} - \mathbf{F})^{1/2}\Lambda^{1/2}\mathbf{B}^{1/2} \right\|_F^2 \geq 0. \end{aligned}$$

In the above, $\|\cdot\|_F$ is the Frobenius norm of a matrix. \square

Lemma 2. Suppose $\mathbf{t} \in \mathcal{T}^p$, $\tilde{\mathbf{t}} \in \mathcal{T}^{p+c}$ for some fixed integer $c > 0$ and \mathcal{T} is a bounded interval. Define $\mathbf{F}(\mathbf{t}) = \Lambda^{1/2}\Phi(\mathbf{t})\{\Phi(\mathbf{t})\Lambda\Phi(\mathbf{t})' + \sigma_\epsilon^2\mathbf{I}_p\}^{-1}\Phi(\mathbf{t})\Lambda^{1/2}$. Then if $\mathbf{t} \subseteq \tilde{\mathbf{t}}$, $\mathbf{F}(\tilde{\mathbf{t}}) - \mathbf{F}(\mathbf{t})$ is positive semidefinite.

Proof. Without loss of generality we assume that $\sigma_\epsilon^2 = 1$. Let $\mathbf{A} = \Phi(\tilde{\mathbf{t}})\Lambda^{1/2}$ and $\mathbf{A}_1 = \Phi(\mathbf{t})\Lambda^{1/2}$. Then \mathbf{A} can be partitioned as $\mathbf{A} = (\mathbf{A}'_1, \mathbf{A}'_2)'$ (after some proper permutation of the index in $\tilde{\mathbf{t}}$), where \mathbf{A}_1 and \mathbf{A}_2 are of dimensions $p \times K$ and $c \times K$, respectively. It follows that

$$\mathbf{A}\mathbf{A}' + \mathbf{I}_p = \begin{pmatrix} \mathbf{A}_1\mathbf{A}'_1 + \mathbf{I}_p & \mathbf{A}_1\mathbf{A}'_2 \\ \mathbf{A}_2\mathbf{A}'_1 & \mathbf{A}_2\mathbf{A}'_2 + \mathbf{I}_c \end{pmatrix}.$$

Now let $\mathbf{A}_{11} = \mathbf{A}_1\mathbf{A}'_1 + \mathbf{I}_p$, $\mathbf{A}_{12} = \mathbf{A}'_2\mathbf{A}_1$ and $\mathbf{A}_{22} = \mathbf{A}_2\mathbf{A}'_2 + \mathbf{I}_c$. Then

$$(\mathbf{A}\mathbf{A}' + \mathbf{I}_p)^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22}^{-1} \end{pmatrix},$$

where $\mathbf{A}_{22}^{-1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. Note that (i) $\mathbf{F}(\mathbf{t}) = \mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_1$ and

$$\begin{aligned} \text{(ii) } \mathbf{F}(\tilde{\mathbf{t}}) &= \mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{I}_p)^{-1}\mathbf{A} = (\mathbf{A}'_1, \mathbf{A}'_2)(\mathbf{A}\mathbf{A}' + \mathbf{I}_p)^{-1}(\mathbf{A}'_1, \mathbf{A}'_2)' \\ &= \mathbf{A}'_1(\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1})\mathbf{A}_1 - \mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_2 - \\ &\quad (\mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_2)' + \mathbf{A}'_2\mathbf{A}_{22}^{-1}\mathbf{A}_2. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbf{F}(\tilde{\mathbf{t}}) - \mathbf{F}(\mathbf{t}) &= \mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_1 - \mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_2 - \\ &\quad (\mathbf{A}'_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_2)' + \mathbf{A}'_2\mathbf{A}_{22}^{-1}\mathbf{A}_2 \\ &= \mathbf{B}'_1\mathbf{B}_1 - \mathbf{B}'_1\mathbf{B}_2 - \mathbf{B}'_2\mathbf{B}_1 + \mathbf{B}_2\mathbf{B}'_2 \\ &= (\mathbf{B}_1 - \mathbf{B}_2)'(\mathbf{B}_1 - \mathbf{B}_2), \end{aligned}$$

where $\mathbf{B}_1 = \mathbf{A}_{22}^{-1/2}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_1$ and $\mathbf{B}_2 = \mathbf{A}_{22}^{-1/2}\mathbf{A}_2$. Therefore, we derive that

$$\mathbf{F}(\tilde{\mathbf{t}}) - \mathbf{F}(\mathbf{t}) = (\mathbf{B}_1 - \mathbf{B}_2)'(\mathbf{B}_1 - \mathbf{B}_2),$$

which is always positive semidefinite. \square

Proof of Theorem 2. In light of Assumption 1 and Eq. (7), it suffices to prove the theorem for the design objective function in (7) with an arbitrary positive semidefinite matrix \mathbf{B} of finite operator norm.

We first focus on the fixed design. Without loss of generality we assume that $\sigma_\epsilon^2 = 1$. Let $\mathbf{U}\mathbf{D}\mathbf{U}'$ be the eigendecomposition of $\Phi(\mathbf{t}_p)\Lambda\Phi(\mathbf{t}_p)'$. Let $\mathbf{V} = \mathbf{U}'\Phi(\mathbf{t}_p)\Lambda^{1/2}$. Note that we drop the dependence of \mathbf{U} and \mathbf{V} on p for simplicity. We also simply denote $\Phi(\mathbf{t}_p)$ by Φ . Then $\mathbf{V}\mathbf{V}' = \mathbf{D}$ and $\Phi(\mathbf{t}_p)\Lambda^{1/2} = \mathbf{U}\mathbf{V}$. It follows that

$$\Lambda^{1/2}\Phi'\Sigma^{-1}\Phi\Lambda^{1/2} = (\mathbf{U}\mathbf{V})'(\mathbf{U}\mathbf{D}\mathbf{U}' + \mathbf{I}_p)^{-1}(\mathbf{U}\mathbf{V}) = \mathbf{V}'(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V}.$$

Then we derive that

$$\mathcal{M}(\mathbf{t}) = \text{tr}\left[\Lambda^{1/2}\mathbf{B}\Lambda^{1/2}(\mathbf{I} - \mathbf{V}'(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V})\right] \leq \|\mathbf{B}\|_{op}\text{tr}\left[\Lambda(\mathbf{I} - \mathbf{V}'(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V})\right],$$

where $\|\mathbf{B}\|_{op}$ denotes the operator norm of \mathbf{B} . Therefore, it suffices to prove that as $p \rightarrow \infty$, $\text{tr}\left[\Lambda(\mathbf{I} - \mathbf{V}'(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V})\right] \rightarrow 0$.

Denote the j th diagonal of \mathbf{D} by d_j . Let $\mathbf{V} = (v_{jk})$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$. Since $\mathbf{V} = \mathbf{U}'\Phi(\mathbf{t}_p)\Lambda^{1/2}$, we obtain $v_{jk} = \mathbf{u}'_j\phi_k\lambda_k^{1/2}$, where $\phi_k = \{\phi_k(t_1), \phi_k(t_2), \dots, \phi_k(t_p)\}'$. Fix an arbitrary small ϵ with $0 < \epsilon < 1/2$. Then there exists an integer K_0 such that $\sum_{k>K_0}\lambda_k < \epsilon$. By Proposition 5.1 in Bunea and Xiao (2015), for any sufficiently large p (depending on ϵ),

$|v_{kk}^2/(p\lambda_k) - 1| \leq \epsilon$, $|d_k - p\lambda_k| \leq \epsilon$ for all $k \leq K_0$. It follows that

$$\begin{aligned} \text{tr} [\Lambda \{\mathbf{I} - \mathbf{V}(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V}\}] &= \sum_{k \geq 1} \lambda_k \left(1 - \sum_{j \geq 1} \frac{v_{jk}^2}{d_j + 1} \right) \\ &\leq \sum_{1 \leq k \leq K_0} \lambda_k \left(1 - \frac{v_{kk}^2}{p\lambda_k} \frac{p\lambda_k}{d_k + 1} \right) + \sum_{k > K_0} \lambda_k \\ &\leq \sum_{1 \leq k \leq K_0} \lambda_k \left\{ 1 - \frac{p\lambda_k}{p\lambda_k + \epsilon + 1} (1 - \epsilon) \right\} + \epsilon \\ &\leq \sum_{1 \leq k \leq K_0} \lambda_k \left(\frac{1 - \epsilon^2}{p\lambda_k + \epsilon + 1} + \epsilon \right) + \epsilon \\ &\leq \epsilon \sum_{1 \leq k \leq K_0} \lambda_k + \sum_{1 \leq k \leq K_0} \frac{\lambda_k}{p\lambda_k + 1 + \epsilon} + \epsilon \\ &\leq \epsilon \sum_{1 \leq k \leq K_0} \lambda_k + K_0/p + \epsilon. \end{aligned}$$

Therefore, with a sufficient large p , we can make $\text{tr} [\Lambda \{\mathbf{I} - \mathbf{V}(\mathbf{D} + \mathbf{I})^{-1}\mathbf{V}\}]$ smaller than ϵ , up to a finite multiplicative constant. We have proved that $\lim_{p \rightarrow \infty} \mathcal{M}(\mathbf{t}_p) = 0$ for the fixed design.

For the random design, the key is to establish that

$$|v_{kk}^2/(p\lambda_k) - 1| = o_{\mathbb{P}}(1), \quad |d_k - p\lambda_k| = o_{\mathbb{P}}(1) \text{ for all } k \leq K_0,$$

which holds by a proposition (similar to Proposition 5.1 in Bunea and Xiao (2015)) in the Supplementary materials. \square

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csda.2018.01.009>.

References

- Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs, with SAS. Oxford, New York.
- Bohorquez, M., Giraldo, R., Mateu, J., 2015. Optimal sampling for spatial prediction of functional data. *Stat. Methods Appl.* 25, 39–54.
- Bunea, F., Xiao, L., 2015. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to *fpca*. *Bernoulli* 21 (2), 1200–1230.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2016. shiny: Web Application Framework for R. R package version 0.13.2.
- Crainiceanu, C., Staicu, A., Ray, S., Punjabi, N., 2012. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Stat. Med.* 31, 3223–3240.
- Delaigle, A., Hall, P., Bathia, N., 2012. Componentwise classification and clustering of functional data. *Biometrika* 99, 299–313.
- Ferraty, F., Hall, P., Vieu, P., 2010. Most-predictive design points for functional data predictors. *Biometrika* 97 (4), 807–824.
- Goldsmith, J., Kitago, T., 2016. Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 65 (2), 215–236.
- Greven, S., Crainiceanu, C., Caffo, B., Reich, D., 2010. Longitudinal functional principal component analysis. *Electron. J. Stat.* 4, 1022–1054.
- Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer, New York.
- James, G., Hastie, T., Sugar, C., 2000. Principal component models for sparse functional data. *Biometrika* 87, 587–602.
- Ji, H., Müller, H.-G., 2017. Optimal designs for longitudinal and functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 79 (3), 859–876.
- Jiang, C.-R., Aston, J.A., Wang, J.-L., 2009. Smoothing dynamic positron emission tomography time courses using functional principal components. *NeuroImage* 47 (1), 184–193.
- Li, H., Staudenmayer, J., Carroll, R.J., 2014. Hierarchical functional data with mixed continuous and binary measurements. *Biometrics* 70 (4), 802–811.
- Lindquist, M.A., 2012. Functional causal mediation analysis with an application to brain connectivity. *J. Amer. Statist. Assoc.* 107 (500), 1297–1309.
- Lu, X., Marron, J.S., 2014. Analysis of spike train data: Comparison between the real and the simulated data. *Electron. J. Stat.* 8 (2), 1793–1796.
- Morris, J., Arroyo, C., Coull, B., Ryan, L., Herrick, R., Gortmaker, S., 2006. Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *J. Amer. Statist. Assoc.* 101, 1352–1364.
- Park, S.Y., Staicu, A.-M., 2015. Longitudinal functional data analysis. *Stat* 4 (1), 212–226.
- Peng, J., Paul, D., 2009. A geometric approach to maximum likelihood estimation of functional principal components from sparse longitudinal data. *J. Comput. Graph. Statist.* 18, 995–1015.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramsay, J., Silverman, B., 2005. Functional Data Analysis. Springer, New York.
- Ramsay, J., Silverman, B.W., 2002. Applied Functional Data analysis: Methods and Case Studies. Springer, New York.
- Randolph, T.W., Harezlak, J., Feng, Z., 2012. Structured penalties for functional linear models - partially empirical eigenvectors for regression. *Electron. J. Stat.* 6, 323–353.
- Rasekhi, M., Jamshidi, B., Rivaz, F., 2014. Optimal location design for prediction of spatial correlated environmental functional data. *J. Mod. Appl. Statist. Meth.* 13 (2), 26.
- Reimherr, M., Nicolae, D., 2014. A functional data analysis approach for genetic association studies. *Ann. Appl. Stat.* 8 (1), 406–429.

- Reiss, P.T., Ogden, R.T., 2010. Functional generalized linear models with images as predictors. *Biometrics* 66 (1), 61–69.
- Ritter, K., 1996. Asymptotic optimality of regular sequence designs. *Ann. Statist.* 24 (5), 2081–2096.
- Tang, R., Müller, H., 2009. Time-synchronized clustering of gene expression trajectories. *Biostatistics* 10, 32–45.
- Wu, M., Diez-Roux, A., Raghunathan, T.E., Sanchez, B.N., 2017. FPCA-based method to select optimal sampling schedules that capture between-subject variability in longitudinal studies. *Biometrics* n/a–n/a.
- Xiao, L., Huang, L., Schrack, J.A., Ferrucci, L., Zipunnikov, V., Crainiceanu, C.M., 2015. Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics* 16 (2), 352–367.
- Xiao, L., Li, C., Checkley, W., Crainiceanu, C., 2017. Fast covariance estimation for sparse functional data. *Stat. Comput.*
- Yao, F., Müller, H., Wang, J., 2005. Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* 100, 577–590.
- Ylvisaker, D., 1987. Prediction and design. *Ann. Statist.* 15 (1), 1–19.